Part 1: Explanation Practice (Homework 11)

Lab 2: What Makes a Product Successful? - W203 Section 8

Team Herkimer: Rick Chen, Chase Madson, Maria Manna, Jash Sompalli

Jul 19, 2022

1 Omitted Variables

Question: *An omitted variable is whether the company has high value assets that might be attractive targets for criminals. Argue whether the omitted variable bias is towards zero or away from zero (5 sentences max).*

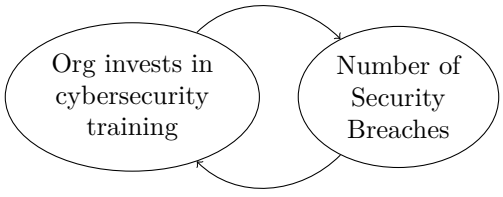
Response: Since the assets of a company are not accounted for in the model, we are likely omitting an important variable that explains the variation around security breaches. It is reasonable to assume criminals are more strongly motivated to breach a company when they believe there are valuable assets to be pilfered, and thus we would expect the number of breaches to be greater for these companies with high value assets. Companies with high value assets are likely the ones investing the most into cybersecurity training, so the coefficient for training hours included in the model is absorbing this positive association.

Thus, the omission of a company's assets is likely producing omitted variable bias in our model, tilting our negative coefficient for training hours towards zero, passing zero, and continuing away from zero in the positive direction. If we were to correct this by including a "High Value Assets" variable, we expect to see a positive and statistically significant coefficient associated with it and it's possible we actually see the coefficient for training hours become negative (i.e., flipping the slope from positive to negative).

2 Reverse Causality

Question: *Explain why there is a possibility of reverse causality. Argue whether the direction of bias is towards zero or away from zero (5 sentences max).*

Response: Companies with high value assets are likely the ones experiencing the most attention from cybercriminals, and so to remedy this they use their greater means to invest in cybersecurity training. This likely means there is reverse causality that a company with higher security breaches tends to cause higher cybersecurity training hours, with a positive feedback to the original coefficient (i.e., $\gamma_1 > 0$ for $Training = \gamma_0 + \gamma_1 Breaches$.



Thus, the original coefficient is likely biased and is greater than the true coefficient, similarly to the omitted variable bias, we expect the bias drags our negative coefficient for true training hours towards zero, passing

zero, and continuing away from zero in the positive direction. Our colleague is misinterpreting a reverse causality in these model results to suggest:

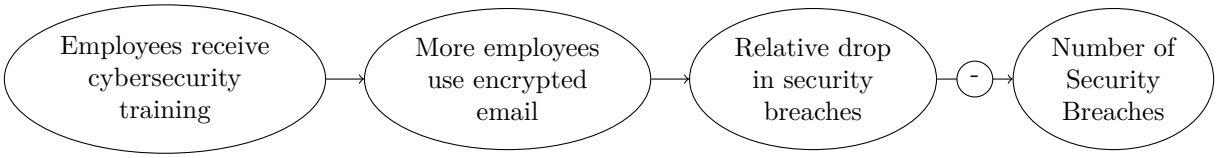
To make matters worse, if the Cybersecurity Training Hours variable is a total number of hours rather than per capita, it could also be inadvertently measuring companies with more employees who are likely to be companies with more assets to begin with.

3 Outcome on Right Hand Side

Question: *Explain why there is an outcome variable on the right hand side. Argue whether removing it would make the coefficient on Cybersecurity Training Hours move up or down (5 sentences max).*

Response: Emails Encrypted is a reflection of how well employees at large comply internalize good cybersecurity practices, which itself is an outcome of having received adequate cybersecurity training. Removing the Emails Encrypted variable will make the coefficient on cybersecurity training hours *move down* towards zero and eventually flip the sign and become negative since the coefficient of Emails Encrypted is currently absorbing much of the mitigating impact Training Hours has on Security Breaches. Much of the impact of having *well-trained* employees is not being counted towards the coefficient for training hours if we are additionally accounting for the best practices directly through Emails Encrypted.

We can imagine Emails Encrypted as a intermediary step we add to the above causal chart between training and relative drop in breaches:



4 Conclusion

Question: *Provide a one-sentence conclusion addressing the idea that your company should not invest in cybersecurity training.*

Response: Let's not get our causes and outcomes mixed up due to this poorly specified regression model - cybersecurity training is an effective and evidence-based method for reducing costly breaches and we would be unwise to forgo it, especially considering the density of high value assets at our company.